DOCUMENT RESUME

ED 381 570 TM 022 930

TITLE	College-Level Academic Skills Test, Technical Report, 1993-94.
INSTITUTION	Florida State Dept. of Education, Tallahassee.
PUB DATE	94
NOTE	46p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; Accountability; Achievement Tests; Associate Degrees; *College Students; *Communication Skills; *Degree Requirements; Higher Education; Language Skills; *Mathematics Achievement; Public Colleges; Racial Differences; Reading Achievement; Scores; Scoring; Sex Differences; State Legislation; Test Construction; *Test Results
IDENTIFIERS	*College Level Academic Skills Test; *Florida

ABSTRACT

The College-Level Academic Skills Test (CLAST), mandated by state law, is part of Florida's system of educational accountability. The CLAST is an achievement test measuring students' college-level communication and mathematics skills as identified by faculties of community colleges and state universities. Since August 1, 1984 students in public institutions in Florida have been required to have CLAST scores that satisfy state standards for the award of an associate in arts degree and for admission to upper division status in a state university in Florida. In addition, students in private schools may need CLAST scores to receive state financial aid. The CLAST consists of essay, English language skills, reading, and mathematics tests. Test development is traced, and the test itself is described, along with scoring and administration information. Summary data are presented for first-time takers in 1993-94 classified by race/ethnicity and gender, as well as college status. Fifty-three percent of students passed the CLAST in June 1994. Fourteen tables present test data for the 1993-94 school year. Six appendixes describe the test in greater detail and list College-Level Academic Skills Project (CLASP) and state-level task force members, 1993-94. (Contains 15 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document *
****

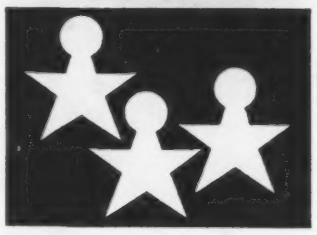
U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
ROBERT BEDFORD
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 381 570

CLAST

COLLEGE-LEVEL ACADEMIC SKILLS TEST



TECHNICAL REPORT

1993-94

Copyright
State of Florida
Department of State
1994

Authorization is hereby granted to elements of the Florida State System of Educational Accountability, as defined in Chapter 229 of the Florida Statutes, to reproduce for use in Florida state and non-public postsecondary institutions.

BEST COPY AVAILABLE

TM022930

TABLE OF CONTENTS

I. OVERVIEW 1

Eligibility to Take the CLAST 1
Test Administration Plan 2

II. DEVELOPMENT OF THE CLAST 3

Background 3
Identification of Skills 3
Review of Available Tests 4
Development of Test Specifications 5
Development of Item Specifications 5
Development of Items 5
Development of CLAST Standards 6

III. DESCRIPTION OF THE CLAST 9

Test Specifications 9
Item Bank 9
Test Assembly 9
Test Instructions 10
Quality Control 11

IV. TECHNICAL CHARACTERISTICS OF THE CLAST 13

Test Score Equating 13
Reliability of Scores 15
Item Analysis 17
Preventing Item Bias 19
Validity of Scores 19

V. SCORING AND REPORTING PROCEDURES 21

Scoring Activities 21
Score Scales 21
Essay Scoring 25
Reporting Test Results 27
Interpreting and Using Scores 29

VI. SUMMARY OF 1993-94 RESULTS 31

BIBLIOGRAPHY 35

APPENDICES 37

TABLES

1. Communication and Computation Competencies and Broad Skills 4
2. Standards (Passing Scores) for CLAST Subtests 7
3. CLAST Specifications by Subtest, 1993-94 10
4. Multiple-Choice Raw Score Reliability Statistics, 1993-94 15
5. Summary Data for All Essay Readers, 1993-94 16
6. Essay Reader Agreement after Referee, 1993-94 16
7. Alpha Coefficients, 1993-94 17
8. English Language Skills Score Conversions, 1993-94 22
9. Reading Score Conversions, 1993-94 23
10. Mathematics Score Conversions, 1993-94 24
11. Raw and Scaled Scores, 1993-94 32
12. Percentage of Examinees Passing All Four Subtests, 1993-94 32
13. Essay Mean Scaled Scores, 1993-94 33
14. English Language Skills Mean Scaled Scores, 1993-94 33
15. Reading Mean Scaled Scores, 1993-94 34
16. Mathematics Mean Scaled Scores, 1993-94 34

I. OVERVIEW

The College-Level Academic Skills Test (CLAST) is part of Florida's system of educational accountability and is mandated by Section 229.551(3)(k), FS. The CLAST is an achievement test measuring students' attainment of college-level communication and mathematics skills identified by faculties of community colleges and state universities through the College-Level Academic Skills Project (CLASP). The skills (Appendix A) have been adopted by the State Board of Education (SBE) through Rule 6A-10.0310, FAC. Provisions for keeping the skills list current, maintaining active participation of faculty members in the implementation of the testing program, and administering the test are provided in the *CLAST Test Administration Plan*.

The CLAST consists of four subtests: Essay, English Language Skills (measuring objective writing skills), Reading, and Mathematics. Each subtest yields a single score reported to the student and to the institution needing the scores. Students also receive broad skill information useful in identifying areas of possible strength or weakness. While the CLAST does not yield the skill-by-skill information necessary for full diagnosis of individual student needs, institutions can identify areas of need for groups of students by aggregating scores into broad skills over several administrations. Although CLAST scores relate positively to other measures of academic performance, they do not predict examinees' future performance in upper division programs.

Since August 1, 1984, students in public institutions in Florida have been required to have CLAST scores that satisfy the standards set forth in SBE Rule 6A-10.0312, FAC, for the award of an associate in arts degree and for admission to upper division status in a state university in Florida. In addition, students in private institutions may need CLAST scores to receive state financial aid.

Statutes and rules pertaining to the CLAST requirement are contained in the *CLAST Test Administration Plan*.

■ *Eligibility to Take the CLAST*

The CLAST may be taken by any student who seeks an associate in arts or a baccalaureate degree, has at least eighteen credit hours, and applies to take the test by the deadline established for registration. Students who have previously taken the CLAST and have not passed all subtests may apply at any regular administration to retake the subtest(s) not passed.

In addition, participating colleges and universities are to register other students who meet either of the following criteria:

1. The students are eligible to participate in a State of Florida financial aid program governed by SBE Rule 6A-20.005, FAC.

2. The students are required under provisions of SBE Rule 6A-20.005, FAC, to have CLAST scores to continue their eligibility beyond the academic term in which they register for the CLAST.

Although CLAST scores are not needed to receive an associate in science degree, students who are in that program may be registered for the CLAST if they satisfy the requirements for (1) the associate in arts degree or (2) admission to upper division status.

In all cases, registration of students for the CLAST must be made in an institution that can determine the eligibility of applicants to take the test. Thus, registration normally will be done by the institution in which students are enrolled during the term in which they will take the test. However, an applicant for upper division status at

a state university who needs CLAST scores and meets other eligibility requirements but is not enrolled in an institution that administers the CLAST may be registered for the test in the institution that needs the scores.

Students must apply to take the test on or before the registration deadline established for that administration. Students may not retake any subtest for which they already have a passing score. Students may not retake any subtest prior to thirty days from the previous administration of the subtest.

■ *Test Administration Plan*

Under provisions of Section 229.551(3)(k), Florida Statutes, the Commissioner of Education maintains statewide responsibility for the administration of the CLAST.

A plan for the administration of the CLAST for the 1993-94 academic year was issued by the Commissioner in July 1993. The plan, developed by the Department of Education, assigns administrative responsibility for the CLAST at three levels: the Department of Education; the Statewide Test Administrator (a technical support contractor); and the community colleges and state universities which administer the test to eligible students. The Office of Instructional Resources of the University of Florida is the Statewide Test Administrator.

The plan also describes the policies and procedures under which the testing program operates. The *CLAST Test Administration Manual* and the *CLAST Institutional Test Administrator's Manual*, which are made a part of the plan, give additional specific information to assist institutional personnel in carrying out their responsibilities.

II. DEVELOPMENT OF THE CLAST

The test development process for the CLAST began with identifying skills to be assessed and continues with developing items for inclusion in the test. This chapter describes the major developmental efforts culminating in the first test administration, the item development procedures, and the development of standards (passing scores).

■ *Background*

In 1979 the Florida Legislature, through Florida Statute 79-222 (now Section 229.551), enacted legislation requiring the identification of skills to measure the achievement of essential academic skills of college students. The Department of Education then charged the Articulation Coordinating Committee with the task of implementing that part of the legislation dealing with the identification of skills and tests to measure achievement of those skills. The result was the establishment of the Essential Academic Skills Project (EASP, now CLASP). The EASP included an executive committee, a project director, a state-level task force on communication, a state-level task force on computation, and a state-level standing committee on student achievement. Members of these initial groups are identified in the *CLAST Technical Report, 1982-83*; current members are identified in Appendix B.

■ *Identification of Skills*

The state-level task forces, together with the project director and other project personnel acting in an advisory capacity, worked to identify essential academic skills that every student, regardless of major, should have acquired by the end of the sophomore year. The task forces worked through a series of meetings from January to November of 1980 with input from institutional-level task forces that had been established to involve faculty members in Florida's public universities and community colleges in the identification of the skills.

The task forces identified four generic competencies (reading, listening, writing, and speaking) in communication and four generic competencies (algorithms, concepts, generalizations, and problem solving) in computation. Each generic competency was subsequently reviewed and broad skill categories were developed for each competency.

Skills were then developed for each broad skill category. These skills were presented to a random sample of faculty members from broad discipline areas in Florida's public community colleges and universities. Based on the results of the survey, the task forces made recommendations to the SBE. In September 1981 the SBE adopted all of the skills recommended by the task forces. During 1985 and 1989, an extensive review of the CLASP skills resulted in the addition, deletion, and/or modification of some of the original skills. As a result of the 1985 review, revised skills were adopted by the SBE and have been measured by the CLAST since the fall 1987 administration (see Table 1); the revised skills resulting from the 1989 review will be incorporated into the CLAST's fall 1992 administration.

TABLE 1
Communication and Computation Competencies and Broad Skills

COMMUNICATION	COMPUTATION
READING Literal Comprehension Critical Comprehension LISTENING Literal Comprehension Critical Comprehension WRITING Multiple-Choice Word Choice Sentence Structure Grammar, Spelling, Capitalization, and Punctuation Essay Suitability to Purpose and Audience Effectiveness and Conformity to Standard English SPEAKING Composition of Message Transmission of Message	ALGORITHMS Arithmetic Geometry and Measurement Algebra Statistics, including Probability Logical Reasoning CONCEPTS Arithmetic Geometry and Measurement Algebra Statistics, including Probability Logical Reasoning GENERALIZATION Arithmetic Geometry and Measurement Algebra Statistics, including Probability Logical Reasoning PROBLEM SOLVING Arithmetic Geometry and Measurement Algebra Statistics, including Probability Logical Reasoning

■ *Review of Available Tests*

Once the skills had been identified, the Standing Committee on Student Achievement, with the assistance of project staff, began its task of identifying tests and other assessment procedures that could be used to measure achievement of the skills. To accomplish the task, an extensive search was conducted to review commercially available tests and tests developed by community colleges and state universities which might be appropriate for measuring achievement of communication and computation skills. Sixty-six communication tests and fifty-four computation tests were reviewed in depth. Though all of the tests addressed some of the skills, none was judged adequate for measuring all of the skills identified in SBE Rule 6A-10.0310, FAC.

It was recommended that three multiple-choice subtests be developed in the areas of writing, reading, and computation. Since all of the writing skills could not be tested using a multiple-choice format, it was further recommended that an essay test be developed to measure the entire set of writing skills. Although it was determined that the identified listening and speaking skills should be acquired by students upon completion of their sophomore year, no statewide tests were developed to measure student achievement of those skills.

A more detailed report on the test search may be found in *Test Search and Screen for College-Level Communication and Computation Skills* (Department of Education, May 1981).

■ *Development of Test Specifications*

Specifications for a test that could be used to measure the achievement of the skills listed in SBE Rule 6A-10.0310, FAC, were developed between April and August of 1981 by the project director and staff, with assistance from the Standing Committee on Student Achievement, the communication and computation task forces, and measurement consultants. Recommendations of state-level task force members about the assessment of the skills, as well as practical and measurement issues, were considered in determining the nature of the subtests and the number of items to be included in each subtest. These same procedures were followed for revising the test specifications necessitated by the 1985 and 1989 skill revisions. Specifications for the 1993-94 forms are described in Chapter III.

■ *Development of Item Specifications*

After test specifications were developed, formulation of item specifications began. During the fall of 1981, item specifications were written for the reading and writing skills as well as for the computation skills dealing with algorithms and concepts. In 1983, item specifications for computation skills dealing with generalizations and problem solving were written and reviewed. Concurrently, the original specifications for the essay, writing, and reading items were reviewed again and revised as necessary. This process was repeated following the 1985 and 1989 skill revisions.

All specifications were written by the chairpersons of the state-level task forces with assistance from task force members, standing committee members, content and measurement consultants, and Department of Education staff. Reviews of the specifications were conducted by faculty members from community colleges and state universities.

Item writers used the item specifications as guides for item content and format. Copies of item specifications were distributed for use in all thirty-seven community colleges and state universities to aid faculties in planning for instruction and assessment of the skills. Copies of item specifications are available in the institutions as well as from the Department of Education.

■ *Development of Items*

Items are developed for the CLAST through contracts with postsecondary faculty who write, review, pilot-test, and revise items based on item specifications and recommendations of state-level item review committees. Items developed under these contracts are submitted to the Department of Education for field-testing and analysis. The following procedures are used to develop and approve test items for the CLAST.

1. A contractor is selected based on its qualifications, including its past performance as an item developer and the qualifications of its item writers and reviewers.

2. The contractor holds a training session for item writers and reviewers to discuss test security issues, the purpose of the CLAST, the use of item specifications, characteristics of good test items, item bias issues, and specific assignments to the contractor.

3. Initial drafts of items are written and reviewed by members of the contractor's item writing team.

4. Items are pilot-tested with college students, and the results of the pilot test and suggestions from other item writers are used in revising the items. The pilot test involves administering each item to about thirty students and interviewing at least five of them to obtain specific information about the items.

5. Based on pilot-test data, items are reviewed and revised by members of the contractor's review team who have not been involved in the item writing. Attention is given to content, measurement, and bias issues (Appendix C).

6. Revised items are submitted to the Department of Education, and a state-level committee is convened to review the items and recommend revisions and/or deletions in the contractor's set.

7. Based on state-level review, items are revised by the contractor's team and submitted to the Department of Education in final form.

8. Items are then included in the CLAST as developmental items and are not counted as scored items for students. This produces classical and Rasch item statistics for evaluating item quality. Items are screened based on the following criteria: p-value greater than or equal to .40, point-biserial greater than or equal to .30, Rasch fit between less than or equal to 3.0, and Rasch total fit less than or equal to 1.0 + 3 standard errors. These criteria represent an ideal level of functioning for an item. If the item point-biserial statistic is less than 0.30, the item may still be considered for use on a future examination if it measures an important dimension of a required objective. Items are not used if the point biserial correlation coefficients are close to or less than zero.

9. Essay topics are field-tested by a qualified contractor. Data generated for topic evaluations include distribution of scores, number of essays written, number written off topic, mean score, median score, percentage of complete agreement between raters, percentage of agreement within one score point, alpha coefficients with and without referee, and reader comments. Topics are evaluated in terms of clarity, relevance and appeal to the target population, and suitability for development of an essay; topics are also screened for potential biasing elements. The contractor recommends the topics suitable for inclusion in the CLAST and identifies any potential problems.

New items for the CLAST Reading, English Language Skills (ELS), and Mathematics subtests were developed in March 1992 through a contract with Florida State University. New topics for the Essay subtest were developed in 1991.

In the fall 1992 administration, changes in the ELS test format resulting from the 1989 skills review and 1990 revision of item specifications were implemented. (Most ELS skills are now tested with items tied to specific passages.) The October 1992 Reading and Mathematics subtests also tested new skills resulting from the 1989 skills review, although for these subtests there was comparatively little change in format.

■ *Development of CLAST Standards*

CLAST standards (passing scores) were set by the SBE in March 1984. The passing scores reflected the judgment of a state-level panel of interested persons concerning the minimum level of performance acceptable for the successful completion of the sophomore year in community colleges and state universities in Florida. SBE Rule 6A-10.0312(1), FAC, establishes minimum standards, in terms of scaled scores, for each CLAST subtest for specified periods of time (Table 2). Students are required to meet the set of standards in effect when they first take the CLAST.

TABLE 2
Standards (Passing Scores) for CLAST Subtests

TIME PERIOD	SCALED SCORES			
	ESSAY	ENGLISH LANGUAGE SKILLS	READING	MATHEMATICS
8/1/84 - 7/31/86	4	265	260	260
8/1/86 - 7/31/89	4	270	270	275
8/1/89 - 9/30/91	4	295	295	285
10/1/91 - 9/30/92	5*	295	295	290
10/1/92 and thereafter	6	295	295	295

* Established with a revision of the scoring scale; equivalent to a total score of 4 on the prior scale.

These tiers of standards are viewed by state-level panel members as reasonable expectations for all students, given the instructional program available to students taking the CLAST during each time period. The *CLAST Technical Report, 1983-84* provides a full description of the process through which the standards were developed.

BEST COPY AVAILABLE

III. DESCRIPTION OF THE CLAST

Each form of the CLAST is developed according to specific guidelines which ensure that test forms from one administration to another are parallel in content and that administration procedures are standardized. This chapter describes the guidelines.

■ *Test Specifications*

For each of the three annual administrations (fall, spring, and summer), a different test is created; however, each test measures the same number of items in each broad skill area (Table 3). To increase test security, two forms of each test are printed for each administration. Both forms contain the same scored items, but the order of item placement is different in each form. Developmental items are embedded in each test form in order to collect data needed to add items to the item bank.

The CLAST comprises four subtests. The Essay subtest is presented in a four-page folder; the English Language Skills and Reading subtests are in the same test book, and the Mathematics subtest is in a separate test book.

■ *Item Bank*

As items are developed, they are numbered with a nine-digit code identifying the subtest, skill, sequence number, and graphic. These items are stored in a card file and a word processing file that are updated as items are revised. New items are added to the bank following the review of the developmental items from each administration.

A history and attribute computer file is kept for the item bank and is used in the selection of items for test forms and in the test analysis process. The file includes attributes such as the item code, broad skill code, item flag, date used, and test form. Statistical data include the percentage correct, item point-biserial coefficient, Rasch difficulty, fit statistics, and index of discrimination for each item. Data on items are kept in the active file for six administrations. After that time, a hard copy and a tape record are stored. The computer bank then is rotated to remove the data from the earliest administrations.

■ *Test Assembly*

For each administration, items are drawn from the item bank to meet the test specifications. Items are selected to minimize the difference in difficulty between forms. Current item difficulty values are used in the selection process. Test form item difficulties are centered near zero logits. Small variations in mean difficulty occur, particularly in the reading test where items are tied to specific passages. Alternate forms are adjusted to the common scale by the equating procedures described in Chapter IV.

TABLE 3
CLAST Specifications by Subtest, 1993-94

SUBTEST AND BROAD SKILL	NUMBER OF SKILLS*	NUMBER OF ITEMS		
		SCORED	DEVELOPMENTAL	TOTAL
ESSAY				
(Holistically scored; not tested with objective items.)				
ENGLISH LANGUAGE SKILLS				
Word Choice	2	4		
Sentence Structure	4	12		
Grammar, Spelling, Capitalization, and Punctuation	10	19		
Total	16	35	5	40
READING				
Literal Comprehension	3	9		
Critical Comprehension	10**	27		
Total	13	36	5	41
MATHEMATICS				
Arithmetic	13	12		
Algebra	17	16		
Geometry and Measurement	10	8		
Logical Reasoning	8	6		
Statistics, including Probability	9	8		
Total	57	50	5	55

* Totals may include subskills and alternate skill formats.
** Includes one alternate skill format indicated in item specifications and skill statements.

The plan for format and arrangement of items in test forms is intended to make each form attractive and easy to read. Multiple-choice writing items are grouped by format and content to make the test time-efficient for students.

■ *Test Instructions*

General instructions provided to students contain information about scoring, recording answers, the number of items, and the time allotted for each subtest. Directions state that scores are based on the number of right answers with no correction for guessing.

The CLAST is administered in one session, which requires nearly five hours. Although actual testing time is four hours, additional time is required to check in examinees, code identifying information, distribute and collect materials, read directions for each subtest, and provide a ten-minute restroom break. The Essay test is administered first, and students are allowed 60 minutes to complete it; the English Language Skills and Reading tests are given next, and 80 minutes are allowed for their completion; the Mathematics test is administered last, and students are given 90 minutes to work on it.

Modifications in test format, such as braille, audio cassette, and large-print materials, are available for handicapped students. In addition, the test schedule and administration procedures are modified for handicapped examinees. Details of these modifications are provided in the *CLAST Institutional Test Administrator's Manual.*

■ *Quality Control*

Test form quality is maintained through an extensive review process. Drafts of new test forms are reviewed by staff of the technical support contractor and the Department of Education. After changes in items and corrections are made, there is a thorough review of camera-ready copy, which is followed by a careful review of bluelines. Additional information about the performance of the test is taken from the institutional test administrators' and room supervisors' reports and on-site visits to test centers by Department of Education personnel. These reports provide information about the quality of test booklets, the standardization of test administrations, and the adequacy of allotted test times.

IV. TECHNICAL CHARACTERISTICS OF THE CLAST

To preserve comparability of CLAST scores from one administration to the next, test scores are equated using a base scale. To ensure reliability and validity of the test and test items, many traditional test analysis procedures are used. This section describes the equating process and procedures used to review the reliability and validity of the test.

■ *Test Score Equating*

The Rasch Model

The CLAST scale development is based on the logistic response model of Georg Rasch, presented in *Probabilistic Models for Some Intelligence and Attainment Tests*, 1960. Rasch describes a probabilistic model in which the probability that a person will answer an item correctly is assumed to be based on the ability of a person and the difficulty of the item. These estimates are derived independently and are not related to the particular sample of people or of items. When the assumptions of the model are met, tests of unequal difficulty can be equated.

Rasch model estimates of person ability and item difficulty are obtained using the unconditional maximum likelihood estimation procedure described in Wright, Mead, and Bell, *BICAL: Calibrating Items With the Rasch Model*, 1980. The probability of a score X_{vi} is expressed as

$$P(X_{vi} \mid B_v, \delta_i) = \frac{\exp [X_{vi} (B_v - \delta_i)]}{1 + \exp [B_v - \delta_i]}$$

where X_{vi} = a score, B_v = person ability, and δ_i = item difficulty.

Person ability in logits represents the natural log odds for succeeding on items which define the scale origin. The item difficulty in logits represents the natural log odds for failure on an item by persons with abilities at the scale origin.

One key assumption of the Rasch model is that a test under consideration is unidimensional. That is, it measures only one underlying student cognitive ability. Unfortunately, ability is considered to be "latent" and cannot be seen or measured in a very precise manner. Therefore, it is important to monitor the performance of the test and to conduct studies that will indicate whether the test is likely to be unidimensional. This has been done with the CLAST examination in two studies. The first study was performed in 1984 with the Computation (Mathematics) subtest. The second was done in 1986 with the Reading, Computation, and Writing (English Language Skills) tests. Both studies showed that the use of Rasch techniques is justified.

Calibration of Items

Item difficulties are obtained by calibrating the scored items for each administration. Three systematic random samples of 700 records are drawn. The items are calibrated, and the item difficulty logits are averaged from the three calibration samples. The averaged difficulties are used to adjust the item logits to the October 1982 base scale.

Item history records are kept in a computer file and updated after each administration. The stability of Rasch difficulty, discrimination values, and fit statistics are checked, and items that change values by more than .3 logit are flagged for further inspection. In addition, following each administration, items are re-examined against established item screening criteria.

Newly developed or revised items are embedded within each form of the test and then calibrated and adjusted to the base scale. These items are not counted toward examinees' scores and are not included in the initial calibrations used to develop the score scale. After the score scale is created, each test form is recalibrated with both the new and the scored items to estimate item difficulties of the new items. The scored items serve as a link between the new items in each test form. Item difficulties for the new items are adjusted to the base scale using the linking constant derived from the comparison of the calibration of the scored items to their base item difficulties. For a complete discussion of the method, see Ryan, J., *Equating New Test Forms to an Existing Test*, 1981.

Generation of Ability Estimates

The traditional estimate of achievement level is the raw score obtained from the number of correct answers provided. The Rasch model is used to generate ability estimates corresponding to the traditional test score.

Adjusted item difficulty logits obtained in item calibration become the basis for estimating person abilities. Generation of ability estimates results in a logit ability scale corresponding to the logit difficulty scale of items. Rasch ability logits are derived using the unconditional maximum likelihood estimation procedures of the program ABIL-EST (Ryan, 1981).

The ability estimate corresponding to each raw score between one point and the number of items minus one is calculated. (Perfect or zero scores are not included in Rasch calculations.) The ability logit scale is then centered at the mean for the October 1982 administration and converted to the standard score scale using a linear transformation.

Linking Scaled Scores

Through the use of Rasch methodology, it is possible to place scores from tests of unequal difficulty on the same scale. While the CLAST difficulty is controlled by selecting items having approximately the same average and range of difficulty for each administration, some fluctuation in difficulty may occur in order to use items representing a broad range of content and difficulty. Differences in test form difficulty are controlled by equating.

Tests forms given on two different occasions are equated by using information obtained from a subset of items common to both forms. These common items are known as "anchor items." The performance of the two groups of examinees on the anchor items is used to adjust the measurement scales for the two forms; the measurement scale for the second form is "adjusted" to that of the first form. From a measurement perspective, the examinees in both instances took the same form of the test. For the CLAST, all test forms are equated back to the first administration of October 1982. With this approach, all students face identical hurdles in that no student has the advantage of an "easier" form.

For each administration, CLAST item difficulties have been adjusted to the base scale of October 1982. Item logits obtained from calibrating the scored items are adjusted by adding the linking constant to each item logit. The difference in average difficulty represents the shift in overall difficulty between test forms. This constant is added to the current item logits to adjust them to the base scale. The stability of the link is evaluated by comparing the difficulty values over time to the values in the base scale.

■ *Reliability of Scores*

Reliability is an indicator of the consistency in measurement of student achievement. It provides an estimate of the variation in results that can be attributed to random error in measurement. The index of reliability is interpreted as the ratio of true-score variance to observed-score variance. Reliability is estimated somewhat differently for multiple-choice scores and essay ratings. Procedures used with each type of score are described in the following sections.

Reliability of Multiple-Choice Scores

The reliability of multiple-choice subtest scores is estimated using the Kuder-Richardson Formula 20 (KR-20) coefficient and the standard error of measurement (SEM). The KR-20 coefficient is an internal consistency estimate of reliability, proposed by Kuder and Richardson in 1937, based on the concept that achievement on items drawn from the same content domain should be related. The formula reported as the KR-20 is

$$r_{tt} = \frac{k}{k-1}\left[\frac{s_t^2 - \Sigma\, pq}{s_t^2}\right]$$

where r_{tt} = estimated test reliability, k = number of test items, s_t^2 = variance of examinees' total scores, and $\Sigma\, pq$ = sum of item variances.

The KR-20 coefficient is appropriate for estimating reliability of scores on multiple-choice tests. However, the KR-20 coefficient can be affected by the distribution of scores. For this reason, the SEM is also reported as an indicator of reliability for each multiple-choice subtest.

The SEM represents the expected standard deviation of scores for an individual taking a large number of randomly selected parallel tests. The mean of the set of scores would represent the individual's true score. Therefore, the SEM can be used to estimate confidence intervals around an individual's true score. Confidence intervals applied to obtained scores are not symmetrical about the obtained score, but the estimated true score is useful in obtaining the center for a confidence zone to be used with the obtained score. The smaller the SEM, the less dispersed are the parallel test scores and the more likely the estimate is close to the individual's true score.

The formula for computing the SEM is $SEM = s_t\sqrt{1-r_{tt}}$ where s_t = standard deviation of the test scores and r_{tt} = test reliability coefficient.

The KR-20s and SEMs for the CLAST multiple-choice subtests indicate that they are acceptably reliable (Table 4).

TABLE 4
Multiple-Choice Raw Score Reliability Statistics, 1993-94

	English Language Skills			Reading			Mathematics		
	OCTOBER	FEBRUARY	JUNE	OCTOBER	FEBRUARY	JUNE	OCTOBER	FEBRUARY	JUNE
KR-20	.72	.73	.71	.71	.70	.79	.86	.84	.85
SEM	1.95	1.89	2.06	2.30	2.32	2.31	2.73	2.81	2.86

Reliability of Essay Ratings

Reliability of essay ratings is evaluated in several ways to ensure that raters have adhered to established criteria for scoring essays. Consistency in scoring is maintained by training the raters and monitoring the scoring process; the reliability of the combined ratings is estimated by coefficient alpha. Both procedures are described below.

Training prior to and during scoring is used to develop and maintain consistency in scoring by the individual rater and the group of raters. The scoring process is monitored by checking the assignment of ratings, the number of split ratings, and the distribution of ratings of each reader. All papers assigned non-contiguous ratings are submitted to a referee who resolves the split scores. During and after each reading session, reader agreement data reflecting the reliability of ratings are reviewed. For the 1993-94 test administrations, the percentage of complete agreement between readers for all papers ranged from 41.3 to 44.9, while the percentage of non-contiguous scores ranged from 9.6 to 11.1 (Table 5). The complete agreement, by topic, resulting from the assignment to a referee of papers with non-contiguous scores was between 47 and 49% (Table 6).

TABLE 5
Summary Data for All Essay Readers, 1993-94

	October		February		June	
	NUMBER	PERCENT	NUMBER	PERCENT	NUMBER	PERCENT
Total Papers Read	17,317	100.0	18,792	100.0	10,314	100.0
Non-Contiguous Scores	1,928	11.1	1,808	9.6	1,024	9.9
Total Agreement Between Readers	7,157	41.3	8,286	44.1	4,531	43.9

TABLE 6
Essay Reader Agreement after Referee, 1993-94

	% Complete Agreement			% Agreement within One Point		
	OCTOBER	FEBRUARY	JUNE	OCTOBER	FEBRUARY	JUNE
TOPIC 1	47	49	49	53	51	51
TOPIC 2	47	48	48	53	52	52

Reliability of combined ratings for essays is estimated by coefficient alpha, which gives the expected correlation between combined ratings of the scoring team and those of a hypothetical parallel team doing the same task. The formula is

$$r_{kk} = \frac{k}{k-1}\left[1 - \frac{\Sigma s_i^2}{s_t^2}\right]$$

where r_{kk} = coefficient of reliability, k = number of test items, Σs_i^2 = sum of item variances, and s_t^2 = variance of examinees' total scores.

Alpha coefficients by topic for the ratings from 1993-94 show they are consistent across topics and administrations (Table 7).

TABLE 7 Alpha Coefficients, 1993-94						
	Non-Refereed Scores			Refereed Scores		
	OCTOBER	FEBRUARY	JUNE	OCTOBER	FEBRUARY	JUNE
TOPIC1	.71	.72	.72	.86	.85	.85
TOPIC2	.71	.73	.74	.85	.87	.86

Reliability of Pass/Fail Classification

Since CLAST scores are used to determine whether students in Florida's community colleges and universities have achieved the level of performance required for the award of an associate in arts degree or for admission to upper division status, reliability in testing and retesting is an important issue. The reliability issue of interest is whether students would consistently pass or would consistently fail if several parallel forms of the test were administered to them. The results of a test-retest study conducted in 1984 indicate that the CLAST is reliable for making pass/fail decisions based on the 1984-86 standards. A complete report of the study is available from the Department of Education, and a summary is available in Appendix D.

■ *Item Analysis*

An item analysis such as the one shown in Figure 1 is prepared for the total group of examinees, each gender, and each racial/ethnic category. These analyses include the number and percentage of examinees who chose each item response, who omitted the item, or who gridded more than one response. In addition, they include item difficulty (proportion of examinees choosing the correct response), item discrimination, and point biserial correlation.

Following test administration, preliminary item analyses are run on the first answer sheets received for scoring. Results of these analyses are screened for item flaws or key errors. Clues to such errors are low discrimination indices or Rasch fit statistics with high values. Other indicators of problems include lack of balance in foil distributions or inordinate difficulty. Items exhibiting these characteristics are flagged and, following a Department of Education review, may be excluded from scoring.

Pretesting new items embedded in the test forms is another form of quality control. Before an item is added to the bank, it is pretested as a non-scored item, and its item statistics are reviewed. Items not meeting the item selection criteria are examined to determine if they are adequate measures of the skills. Any item deemed inappropriate is flagged and not used on the CLAST.

READING SCORED TOTAL

ITEM RESPONSES — ITEM RESPONSE FIGURES ARE TOTALS, NOT PERCENTAGES

ITEM NUMBER	A	B	C	D	E	OMIT	MULT	ITEM DIFFICULTY	ITEM DISCRIMINATION	POINT BISERIAL CORRELATION
1	4989+	686	2482	1920	0	10	0	0.49	0.41	0.31
2	749	7806+	807	698	0	26	1	0.77	0.41	0.40
3	1714	759	5761+	1843	0	10	0	0.57	0.42	0.33
4	735	8935+	311	101	0	5	0	0.89	0.22	0.29
5	6994+	1137	1254	678	0	22	2	0.69	0.33	0.28
6	250	9808+	14	9	0	5	1	0.97	0.07	0.23
7	640	7421+	1554	443	0	28	1	0.74	0.34	0.32
8	396	1508	7788+	383	0	12	0	0.77	0.31	0.30
9	201	446	214	9217+	0	9	0	0.91	0.23	0.40
10	1450	299	6329+	1995	0	13	1	0.63	0.42	0.35
11	527	113	3257	6180+	0	10	0	0.61	0.45	0.35
12	332	273	8100+	1372	0	10	0	0.80	0.32	0.33
13	7854+	343	1001	867	0	22	0	0.78	0.28	0.29
14	184	444	7545+	1894	0	19	1	0.75	0.30	0.29
15	1426	1117	6005+	1516	0	23	0	0.60	0.52	0.41
16	843	83	87	9056+	0	17	1	0.90	0.19	0.29
17	8927+	285	462	390	0	23	0	0.89	0.28	0.38
18	756	776	728	7708+	0	119	0	0.76	0.49	0.48
19	906	451	7931+	721	0	78	0	0.79	0.45	0.45
20	8457+	700	613	198	0	119	0	0.84	0.34	0.40
21	1341	7387+	603	627	0	126	3	0.73	0.43	0.39
22	9587+	110	142	233	0	14	1	0.95	0.12	0.30
23	8925+	292	520	337	0	13	0	0.88	0.20	0.28
24	77	138	9801+	51	0	20	0	0.97	0.08	0.29
27	994	9028+	28	11	0	26	0	0.90	0.16	0.25
28	9284+	730	46	9	0	18	0	0.92	0.10	0.20
29	7358+	612	678	1404	0	34	1	0.73	0.49	0.46
30	346	3675	482	5552+	0	32	0	0.55	0.51	0.41
31	8932+	946	99	72	0	38	0	0.89	0.30	0.42
34	2587	976	2366	3994+	0	162	2	0.40	0.46	0.36
36	7853+	275	1511	344	0	104	0	0.78	0.26	0.28
37	7603+	793	930	647	0	113	1	0.75	0.40	0.38
38	495	951	1398	7116+	0	127	0	0.71	0.50	0.44
39	2283	6801+	489	370	0	143	1	0.67	0.50	0.42
40	1601	6939+	667	486	0	392	2	0.69	0.49	0.41
41	498	1476	549	6959+	0	605	0	0.69	0.54	0.47

\+ - INDICATES CORRECT ANSWER
** - INDICATES EVERYONE GIVEN CREDIT
* - INDICATES QUESTION THROWN OUT

Figure 1. Example of an item analysis.

■ *Preventing Item Bias*

In addition to examining item analyses, review panels established at each stage of test development considered the issue of bias in the items. Scatter graphs were examined after each administration to determine if particular items operated differently for various racial or ethnic groups.

A scatter graph (Figure 2) contrasts performance on individual items by racial/ethnic or gender categories. An item difficulty is identified as an outlier if it deviates substantially from the general relationship for the compared groups. Consistent differences in item difficulties may indicate only a difference in the level of achievement among the compared groups, but items that deviate from this general pattern are further examined for content bias that may be related to gender or racial/ethnic background.

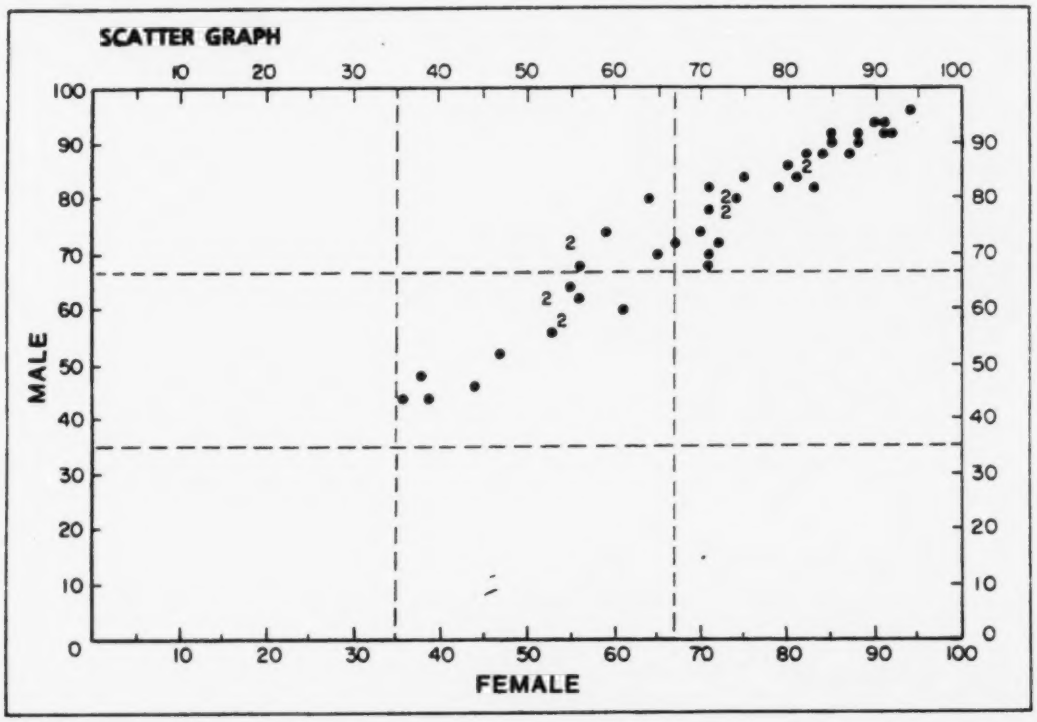


Figure 2. Example of a scatter graph of item difficulties comparing the performance of males with that of females.

■ *Validity of Scores*

Strictly speaking, one should not describe a test as being "valid." Instead, one should describe a test score as being "valid" for a particular purpose. Hence, test development operations are designed to build evidence for a particular type of score interpretation which is defined in advance.

Standards for Educational and Psychological Testing (1985) describes three types of validity: content, construct, and criterion. Content validity is the only important type for the CLAST because test scores are only interpreted in terms of what they indicate about student achievement of designated performance objectives. The CLAST does not measure a designated psychological characteristic (e.g., spatial visualization), so construct validity is not relevant. Further, as has been stated, the CLAST was not designed to predict a student's future performance in

school. Hence, the criterion-related (i.e., predictive) validity is not relevant. Content validity is substantiated by determining the extent to which the test items adequately measure the specific skills they are designed to measure; that is, the extent to which the content of the test matches the set of skills. The validity of the test is established by following the plan and procedures for developing and selecting items for each form of the CLAST.

The general plan used in developing the test is outlined below.

1. General test specifications, consistent with the purpose of the CLAST, are developed by faculty who have expertise in both testing and the content areas (English language skills, reading, and mathematics) with assistance from Department of Education staff.

2. Item specifications, detailing both the content and the format of items that can be developed to measure each of the skills, are developed by faculty with expertise in both the content areas and testing, with assistance from Department of Education staff.

3. Test items are written by faculty according to the guidelines provided by the item specifications and are reviewed by faculty and Department of Education staff with careful attention given to content, measurement, and bias issues.

4. Test items are field-tested in community colleges and state universities.

5. Items are analyzed statistically and selected for use in the test only if they meet criteria established by Department of Education staff and testing consultants.

6. A test plan for selection of items is followed in developing alternate forms of the test.

7. Scaled scores equated to the reference scale are generated using the Rasch model.

To summarize, validity of the test as a measure of achievement of the skills is established by following the plan for developing and selecting items. Content and testing specialists judge the adequacy of the items for measuring the skills, and the plan for selecting items ensures that each form of the CLAST is representative of the domain of skills being tested. Scores on each of the subtests, then, can be interpreted to be valid indicators of students' achievement of the communication and mathematics skills measured by the CLAST.

V. SCORING AND REPORTING PROCEDURES

Procedures for scoring the CLAST are designed to provide quality control and score scale stability for a testing program that has complex scoring and reporting requirements. The process for scoring and reporting reflects concern for reliability and comparability of the scores and for appropriate use of the scores. This chapter addresses those concerns.

■ *Scoring Activities*

Editing Answer Sheets

Following each administration, as answer sheets are received from each institution, they are edited for errors. Answer sheets are read by an NCS Opscan 21 scanner programmed to identify mismarked or miscoded sheets. Each identified answer sheet is hand-checked and corrected according to the scoring conventions.

Rating sheets from holistic scoring of essays are also machine-scored. Editing procedures for holistic scoring include a verification of the legitimacy of reader numbers and score codes. Papers with invalid scores or with ratings that differ by more than one point are returned to the referee to be corrected and/or reviewed.

Scoring Conventions

Within the parameters of number-right scoring, certain conventions are observed: for a response to be considered valid, it must be recorded in the answer folder; for a score to be generated on a subtest, at least one response must be marked in the appropriate section of the answer sheet; and omits and double grids are counted as incorrect. To receive credit for the essay test, students must write on one of the two topics provided, and they must write the essay in their answer folders.

Students' subtest scores below the chance level are compared to their other subtest scores. If a score is inconsistent with the student's performance on the other subtests, it is hand-checked to determine if the student entered the correct form code on the answer sheet.

■ *Score Scales*

A three-digit standard scaled score is generated for each administration for each of the multiple-choice subtests. The standard score scale is a linear transformation of the Rasch ability logits adjusted for the mean of the October 1982 administration. The formula used is

$$S_i = 30(X_b - C) + 300$$

where S_i = scaled score, X_b = ability logit, and C = October 1982 scale adjustment factor (1.87 for English language skills, 1.2 for reading, and 1.0 for mathematics). Raw score to scaled score transformation data are generated for each subtest for each administration (Tables 8, 9, and 10).

TABLE 8

ENGLISH LANGUAGE SKILLS SCORE CONVERSIONS, 1993-94

Raw Score	OCTOBER		FEBRUARY		JUNE	
	Ability	Scaled Score	Ability	Scaled Score	Ability	Scaled Score
0	-6.521	104	-6.386	108	-6.321	110
1	-5.543	133	-5.429	137	-5.357	139
2	-4.798	156	-4.697	159	-4.621	161
3	-4.332	170	-4.246	172	-4.164	175
4	-3.981	180	-3.908	182	-3.822	185
5	-3.693	189	-3.634	190	-3.544	193
6	-3.445	196	-3.399	198	-3.304	200
7	-3.224	203	-3.191	204	-3.092	207
8	-3.024	209	-3.002	209	-2.900	213
9	-2.838	214	-2.829	215	-2.722	218
10	-2.663	220	-2.667	219	-2.556	223
11	-2.497	225	-2.513	224	-2.398	228
12	-2.338	229	-2.367	228	-2.247	232
13	-2.184	234	-2.227	233	-2.101	236
14	-2.035	238	-2.090	237	-1.959	241
15	-1.888	243	-1.957	241	-1.820	245
16	-1.742	247	-1.826	245	-1.682	249
17	-1.598	252	-1.697	249	-1.546	253
18	-1.454	256	-1.569	252	-1.409	257
19	-1.308	260	-1.440	256	-1.272	261
20	-1.161	265	-1.311	260	-1.134	265
21	-1.012	269	-1.180	264	-0.993	270
22	-0.858	274	-1.046	268	-0.849	274
23	-0.700	279	-0.908	272	-0.701	278
24	-0.536	283	-0.766	277	-0.547	283
25	-0.364	289	-0.617	281	-0.386	288
26	-0.182	294	-0.460	286	-0.216	293
27	0.012	300	-0.292	291	-0.034	298
28	0.221	306	-0.109	296	0.162	304
29	0.452	313	0.092	302	0.377	311
30	0.711	321	0.321	309	0.620	318
31	1.011	330	0.588	317	0.901	327
32	1.376	341	0.918	327	1.245	337
33	1.857	355	1.363	340	1.703	351
34	2.622	378	2.088	362	2.440	373
35	3.627	408	3.035	391	3.406	402

TABLE 9

READING SCORE CONVERSIONS, 1993-94

Raw Score	OCTOBER		FEBRUARY		JUNE	
	Ability	Scaled Score	Ability	Scaled Score	Ability	Scaled Score
0	-5.924	122	-6.175	114	-5.883	123
1	-4.958	151	-5.162	145	-4.903	152
2	-4.221	173	-4.391	168	-4.155	175
3	-3.763	187	-3.906	182	-3.690	189
4	-3.421	197	-3.540	193	-3.340	199
5	-3.143	205	-3.242	202	-3.055	208
6	-2.905	212	-2.987	210	-2.811	215
7	-2.695	219	-2.763	217	-2.595	222
8	-2.505	224	-2.561	223	-2.400	228
9	-2.331	230	-2.376	228	-2.220	233
10	-2.168	234	-2.204	233	-2.052	238
11	-2.015	239	-2.043	238	-1.894	243
12	-1.869	243	-1.890	243	-1.744	247
13	-1.729	248	-1.743	247	-1.599	252
14	-1.593	252	-1.602	251	-1.460	256
15	-1.461	256	-1.466	256	-1.324	260
16	-1.332	260	-1.332	260	-1.190	264
17	-1.205	263	-1.201	263	-1.059	268
18	-1.079	267	-1.071	267	-0.929	272
19	-0.953	271	-0.941	271	-0.799	276
20	-0.828	275	-0.812	275	-0.669	279
21	-0.701	278	-0.682	279	-0.537	283
22	-0.572	282	-0.550	283	-0.404	287
23	-0.440	286	-0.416	287	-0.268	291
24	-0.305	290	-0.278	291	-0.129	296
25	-0.165	295	-0.135	295	0.016	300
26	-0.019	299	0.014	300	0.167	305
27	0.136	304	0.171	305	0.327	309
28	0.301	309	0.339	310	0.497	314
29	0.480	314	0.521	315	0.682	320
30	0.678	320	0.721	321	0.886	326
31	0.902	327	0.947	328	1.116	333
32	1.165	334	1.212	336	1.386	341
33	1.489	344	1.539	346	1.717	351
34	1.927	357	1.979	359	2.164	364
35	2.643	379	2.698	380	2.889	386
36	3.578	407	3.637	409	3.837	415

TABLE 10

MATHEMATICS SCORE CONVERSIONS, 1993-94

Raw Score	OCTOBER		FEBRUARY		JUNE	
	Ability	Scaled Score	Ability	Scaled Score	Ability	Scaled Score
0	-6.185	114	-5.842	124	-6.002	119
1	-5.225	143	-4.918	152	-5.066	148
2	-4.490	165	-4.208	173	-4.348	169
3	-4.040	178	-3.779	186	-3.912	182
4	-3.708	188	-3.466	196	-3.591	192
5	-3.441	196	-3.216	203	-3.333	200
6	-3.215	203	-3.005	209	-3.116	206
7	-3.018	209	-2.822	215	-2.925	212
8	-2.842	214	-2.659	220	-2.755	217
9	-2.682	219	-2.510	224	-2.600	222
10	-2.535	223	-2.374	228	-2.457	226
11	-2.397	228	-2.246	232	-2.323	230
12	-2.268	231	-2.127	236	-2.198	234
13	-2.145	235	-2.013	239	-2.078	237
14	-2.029	239	-1.905	242	-1.965	241
15	-1.916	242	-1.801	245	-1.855	244
16	-1.808	245	-1.701	248	-1.750	247
17	-1.703	248	-1.603	251	-1.648	250
18	-1.601	251	-1.508	254	-1.548	253
19	-1.501	254	-1.416	257	-1.451	256
20	-1.403	257	-1.325	260	-1.356	259
21	-1.307	260	-1.235	262	-1.262	262
22	-1.212	263	-1.147	265	-1.170	264
23	-1.117	266	-1.059	268	-1.078	267
24	-1.024	269	-0.972	270	-0.987	270
25	-0.931	272	-0.885	273	-0.897	273
26	-0.837	274	-0.798	276	-0.807	275
27	-0.744	277	-0.711	278	-0.717	278
28	-0.650	280	-0.624	281	-0.626	281
29	-0.556	283	-0.535	283	-0.535	283
30	-0.460	286	-0.445	286	-0.442	286
31	-0.363	289	-0.355	289	-0.349	289
32	-0.264	292	-0.262	292	-0.254	292
33	-0.164	295	-0.167	294	-0.157	295
34	-0.060	298	-0.070	297	-0.057	298
35	0.046	301	0.031	300	0.045	301
36	0.156	304	0.135	304	0.151	304
37	0.271	308	0.243	307	0.261	307
38	0.390	311	0.357	310	0.376	311
39	0.516	315	0.477	314	0.497	314
40	0.650	319	0.604	318	0.625	318
41	0.793	323	0.741	322	0.763	322
42	0.947	328	0.889	326	0.912	327
43	1.117	333	1.052	331	1.076	332
44	1.307	339	1.236	337	1.260	337
45	1.524	345	1.446	343	1.471	344
46	1.781	353	1.696	350	1.721	351
47	2.101	363	2.009	360	2.034	361
48	2.536	376	2.437	373	2.462	373
49	3.253	397	3.146	394	3.172	395
50	4.187	425	4.069	422	4.096	422

The score scale ranges from approximately 100 points to 400 points. It is centered at 300 points, designating the state average score on the October 1982 administration. All subsequent examinations are equated to this administration. Differences in scaled score ranges across test forms occur as a result of differences in the range of item difficulty in test forms. The difficulty of each form is controlled, however, so that these shifts in the average score range are small. If one test form has items that are more difficult, it is possible to obtain a higher scaled score because the harder items measure a higher level of achievement.

The essay score is assigned on a scale of two to twelve points. Two readers rate each essay on a rating scale from one to six points. The essay score is the sum of the two ratings. The holistic scoring procedure and rating scale are discussed in the next section.

■ *Essay Scoring*

Holistic scoring or evaluation, a process for judging the quality of writing samples, has been used for many years by testing agencies in credit-by-examination, state assessment, and teacher certification programs.

Holistic Scores

Essays are scored holistically — that is, for the total, overall impression they make on the reader — rather than analytically, which requires careful analysis of specific features of a piece of writing. Holistic scoring assumes that the skills which make up the ability to write are closely interrelated and that one skill cannot be separated from the others. Thus, the writing is viewed as a total work in which the whole is something more than the sum of the parts. A reader reads a writing sample once, forms an impression of its overall quality, and assigns it a numerical rating based on his/her judgment of how well the paper meets a particular set of established criteria. A six-point scale reflecting the following performance levels is used to score CLAST essays.

Score of 6. The paper presents or implies a thesis that is developed with noticeable coherence. The writer's ideas are usually substantive, sophisticated, and carefully elaborated. The writer's choice of language and structure is precise and purposeful, often to the point of being polished. Control of sentence structure, usage, and mechanics, despite an occasional flaw, contributes to the writer's ability to communicate the purpose.

Score of 5. The paper presents or implies a thesis and provides convincing, specific support. The writer's ideas are usually fresh, mature, and extensively developed. The writer demonstrates a command of language and uses a variety of structures. Control of sentence structure, usage, and mechanics, despite an occasional flaw, contributes to the writer's ability to communicate the purpose.

Score of 4. The paper presents a thesis and often suggests a plan of development, which is usually carried out. The writer provides enough supporting detail to accomplish the purpose of the paper. The writer makes competent use of language and sometimes varies sentence structure. Occasional errors in sentence structure, usage, and mechanics do not interfere with the writer's ability to communicate the purpose.

Score of 3. The paper presents a thesis and often suggests a plan of development, which is usually carried out. The writer provides support that tends toward generalized statements or a listing. In general, the support is neither sufficient nor clear enough to be convincing. Sentence structure tends to be pedestrian and often repetitious. Errors in sentence structure, usage, and mechanics sometimes interfere with the writer's ability to communicate the purpose.

BEST COPY AVAILABLE

Score of 2. The paper usually presents a thesis. The writer provides support that tends to be sketchy and/or illogical. Sentence structure may be simplistic and disjointed. Errors in sentence structure, usage, and mechanics frequently interfere with the writer's ability to communicate the purpose.

Score of 1. The paper generally presents a thesis that is vaguely worded or weakly asserted. Support, if any, tends to be rambling and/or superficial. The writer uses language that often becomes tangled, incoherent, and thus confusing. Errors in sentence structure, usage, and mechanics frequently occur.

Holistic Scoring

The holistic scoring session must be conducted in a highly organized manner with competent staff members who have clearly specified responsibilities. For ten thousand essays, the holistic scoring staff consists of a chief reader, three assistant chief readers, twenty table leaders, and one hundred readers. A support staff of a manager and five clerks is also required.

The scoring procedure follows this pattern. Prior to the scoring session, the chief reader and assistants sample the total group of essays to choose from each of the two topics examples which clearly represent the established standards for each of the six ratings on the rating scale. These essays are known as range finders. In addition, other essays are chosen as training materials during the scoring sessions.

After range finders and samples are selected, table leaders meet with the chief and assistant chief readers to score the samples and determine if the samples clearly represent the six levels of the scale. The purpose of this session is to refine the sample selection and to ensure consensus among table leaders. Range finders from previous administrations are also reviewed and used in the training to ensure consistency in scoring from one administration to another.

Immediately prior to and intermittently throughout the scoring session, the chief reader trains the readers using the range finders and other samples. Immediately after the initial training session, scoring begins. Each essay is read by two readers, each of whom assigns it a rating of one, two, three, four, five, or six. The sum of the ratings is the total score assigned to the essay. A total score of five or above is passing for examinees first taking the CLAST before October 1992.

In situations where the two readers' individual ratings differ by more than one point, the essay is read by a third reader, the referee. The referee's rating will replace one of the existing two ratings for a revised total score.

A more complete description of the process is in *A Training Manual for the Holistic Scoring of the Essay Subtest of the College-Level Academic Skills Test*, available in the Department of Education office.

Recruitment of Readers

Each institution that registers students for the CLAST may participate in the holistic scoring process. The chief reader solicits nominations for readers from the chairs of English departments in community colleges and universities. Nominations for readers are made on the basis of the candidate's interest in the process, willingness to set aside personal standards for judging the quality of writing and to undergo training, and availability to work over weekends. Candidates must have a minimum of two years' experience teaching composition, hold at least a master's degree or equivalent, have a major in English in at least one degree, and teach composition as part of their assigned responsibilities. Nominations may include secondary school teachers who teach composition at the junior or senior year level in high schools and faculty who teach composition in private postsecondary institutions.

Upon receiving nominations from department chairs, the chief reader and the Statewide Test Administrator ask each nominee interested in becoming a reader to complete and submit an application form. The forms are used to determine whether applicants meet the criteria for readers.

■ *Reporting Test Results*

The reports outlined below are generated for each administration. In addition to these reports, institutions may request from the Statewide Test Administrator a computer tape or diskettes containing their students' data, including item responses. Thus, institutions can generate their own reports and update files of students' records. A test blueprint giving item-skill correspondence and a data tape format are also provided to institutions.

Student Reports

The individual student report (Figure 3) and a score interpretation guide are mailed to students approximately five weeks after the examination date. A scaled score is reported for each subtest taken. In the boxes to the right of the scale score is reported the percentage of items correct in each broad skill area. Although the percentages are reported to the student, they do not become part of the student's transcript. The percentages help students determine their relative strengths and weaknesses in the broad skill areas represented on the test, but because of differences in the number of test items used in the broad skill areas, the percentages cannot be averaged to determine the overall percentage of correct responses.

Preliminary Reports--prepared at the state and institutional levels

1. Summary statistics (means, medians, and standard deviations) and frequency distributions of scores by

   a. Student classification:
      Community college A.A. program
      Community college A.S. program
      University native student
      University transfer student

   b. Racial/ethnic classification:
      White/non-Hispanic
      Black/non-Hispanic
      Hispanic
      American Indian/Alaskan native
      Asian/Pacific Islander
      Non-Resident Alien

   c. Gender by racial/ethnic classification

2. Alphabetic roster of examinees' scores

Final Reports--prepared at the state and institutional levels

1. Means and percents of first-time examinees meeting current standards for

   a. students with 60 or more hours
   b. students with fewer than 60 hours

(cont., p. 29)

Individual Score Report
COLLEGE-LEVEL ACADEMIC SKILLS TEST
DATE OF EXAM:

S.S. #

INSTITUTION

Below are your scores on the College-Level Academic Skills Test. In the shaded box on the left is your essay score. In the remaining shaded boxes are your three-digit scale scores for each subtest. After each scale score is printed the following information for each broad skill area of a subtest: number of correct answers number of questions = percentage of items answered correctly. The enclosed interpretation guide will help you understand your scores.

This report is only for your information. The official record of your scores is kept by your institution on your transcript.

ESSAY

Essay Rating

ENGLISH LANGUAGE SKILLS

SCALE SCORE	Word Choice	Sentence Structure	Grammar, Spelling, Punctuation, Capitalization

READING

SCALE SCORE	Comprehension	
	Literal	Critical

MATHEMATICS

SCALE SCORE	Arithmetic	Algebra	Geometry - Measurement	Logical Reasoning	Statistics

Passing scores on CLAST have been established by the State Board of Education as follows:

Essay	English Language Skills	Reading	Mathematics

Students are required to meet the standards in effect at the time they first took the test.

If you have questions about your scores, you should contact:

Figure 3. Copy of a blank student report form.

(from p. 27)

c. state university native students
d. state university transfer students
e. students by gender and racial ethnic category for each institution, all public institutions, all private institutions, all community colleges, and all state universities

2. Means and percents of first-time examinees meeting future standards by gender and racial ethnic category for each institution, all public institutions, and all private institutions

3. Means and percents of retake examinees meeting required standards by gender and racial ethnic category for each institution, all public institutions, and all private institutions

Statistical Reports--prepared at the state level only

1. Rasch item calibrations and fit statistics
2. Scaled score derivations
3. Classical item analysis by racial/ethnic classification
4. Item difficulty plots by gender and racial/ethnic classification
5. KR-20 coefficients and SEM's for multiple-choice subtests
6. Interrater reliability for essay scores
7. Coefficient alpha by gender and racial/ethnic classification for essay scores

■ *Interpreting and Using Scores*

CLAST scores are reported to indicate students' achievement of those skills upon which the test is based. The CLAST scaled scores, not the raw scores, for each subtest are used for this purpose since the scaled scores have been adjusted for differences in difficulty in test forms. A scaled score of 300, for instance, represents the same achievement level across forms but may require a higher raw score on an easier form than on a harder one. The same scaled score, then, represents the same level of achievement of the skills regardless of the test form taken.

The use of CLAST scores is prescribed by Florida Statutes and Rules of the SBE. Use of scores prior to August 1, 1984, was limited to student advising and curriculum improvement. Since August 1, 1984, students in public institutions in Florida are required to have CLAST scores which satisfy the standards set forth in Rule 6A-10.0312, FAC, for the award of an associate in arts degree and for the admission to upper division status in a state university in Florida. However, students who have satisfied CLAST standards on three of the four subtests and who are otherwise eligible may be enrolled in state universities for up to an additional thirty-six semester credits of upper division coursework before they are required to pass the fourth subtest.

Standards (passing scores) for the CLAST have been adopted by the SBE in Rule 6A-10.0312(1), FAC. The standards for each designated period of time are indicated in Chapter II.

The CLAST was not developed to predict success in upper division programs, but to assess the level of achievement of the skills listed in Appendix A. Any use of the scores for selection of students for specific upper division programs must be empirically validated.

VI. SUMMARY OF 1993-94 RESULTS

The results of CLAST administrations indicate the level of achievement of communication and computation skills by students in community colleges and state universities. Summary data presented in this section describe student performance on the CLAST as a whole and on each subtest. Summary data are based on only those students who were first-time takers in public institutions.

The mean, standard deviation, and median of raw scores and scaled scores are reported by subtest for each administration (Table 11). For the Essay, English Language Skills, and Mathematics subtests, mean scaled scores moved in a generally downward trend from the October administration through the February and June administrations. Mean scaled scores on the Reading subtest, however, were highest in the June administration.

The percentage of examinees that passed the CLAST was 59 in October 1993, 57 in February 1994, and 53 in June 1994 (Table 12). The passing rates for groups of students classified on the basis of gender or racial/ethnic background varied across all administrations, ranging from a low of 24% to a high of 75% (Table 12).

Mean scores are reported for all students, for students grouped according to gender, for students grouped according to racial/ethnic background, for students in community colleges, and for students in the state university system. These means are provided separately for the Essay, English Language Skills, Reading, and Mathematics subtests and are found in Tables 13, 14, 15, and 16, respectively.

TABLE 11

Raw and Scaled Scores, 1993-94
(First-Time Examinees in Public Institutions)

	NO. of ITEMS	RAW SCORE			SCALED SCORE		
		MEAN	STD. DEV.	MEDIAN	MEAN	STD. DEV.	MEDIAN
Essay							
October					7.6	1.9	8
February					7.4	1.9	7
June					7.3	1.9	7
English Language Skills							
October	35	29.2	3.7	30	321.3	30.4	321
February	35	30.3	3.6	31	320.7	31.7	317
June	35	28.7	3.8	29	314.6	28.2	311
Reading							
October	36	28.2	4.3	29	313.8	24.5	314
February	36	27.7	4.3	28	312.1	24.2	310
June	36	27.4	5.1	28	316.7	29.8	314
Mathematics							
October	50	37.7	7.2	39	315.7	30.3	315
February	50	37.7	7.1	39	314.4	28.4	314
June	50	36.3	7.4	37	309.5	27.9	307

TABLE 12

Percentage of Examinees Passing All Four Subtests, 1993-94
(First-Time Examinees in Public Institutions)

EXAMINEE GROUP	OCTOBER		FEBRUARY		JUNE	
	Number Tested	Percent Passing	Number Tested	Percent Passing	Number Tested	Percent Passing
All	17,416	59	18,804	57	10,319	53
Male	7,564	60	8,268	59	4,376	54
Female	9,852	59	10,536	55	5,943	52
White	11,945	68	13,150	65	7,065	61
Black	2,162	34	2,015	30	1,023	24
Hispanic	2,212	42	2,402	41	1,552	37
Asian/Pacific Islander	629	50	721	46	392	39
American Indian/Alaskan Native	47	60	55	45	28	75
Non-Resident Alien	304	39	338	40	186	38
Unknown Race	117	44	123	48	73	47
Community College	9,269	52	10,911	51	7,543	50
State University System	8,147	67	7,893	66	2,776	60

BEST COPY AVAILABLE

TABLE 13

Essay Mean Scaled Scores, 1993-94
(First-Time Examinees in Public Institutions)

EXAMINEE GROUP	OCTOBER		FEBRUARY		JUNE	
	Number	Mean	Number	Mean	Number	Mean
All	17,482	7.6	18,913	7.4	10,388	7.3
Male	7,595	7.4	8,309	7.2	4,398	7.1
Female	9,887	7.7	10,604	7.5	5,990	7.4
White	11,999	7.9	13,230	7.7	7,114	7.6
Black	2,171	6.7	2,029	6.7	1,032	6.4
Hispanic	2,213	7.0	2,414	6.8	1,558	6.7
Asian/Pacific Islander	629	7.0	722	6.6	393	6.1
American Indian/AlaskanNative	49	8.2	57	7.4	29	7.4
Non-ResidentAlien	304	6.6	338	6.3	188	6.5
Unknown Race	117	7.1	123	7.1	74	7.0
Community College	9,319	7.4	10,995	7.1	7,604	7.2
State University System	8,163	7.8	7,918	7.7	2,784	7.6

TABLE 14

English Language Skills Mean Scaled Scores, 1993-94
(First-Time Examinees in Public Institutions)

EXAMINEE GROUP	OCTOBER		FEBRUARY		JUNE	
	Number	Mean	Number	Mean	Number	Mean
All	17,501	321	18,910	321	10,392	315
Male	7,611	318	8,306	318	4,402	312
Female	9,890	324	10,604	323	5,990	317
White	12,011	326	13,230	325	7,118	320
Black	2,171	307	2,029	304	1,033	300
Hispanic	2,216	313	2,412	312	1,556	305
Asian/Pacific Islander	630	316	722	314	395	306
American Indian/AlaskanNative	49	323	56	321	29	315
Non-ResidentAlien	307	308	338	309	188	307
Unknown Race	117	315	123	313	73	314
Community College	9,325	317	10,997	316	7,607	313
State University System	8,176	326	7,913	327	2,785	320

BEST COPY AVAILABLE

TABLE 15

Reading Mean Scaled Scores, 1993-94
(First-Time Examinees in Public Institutions)

EXAMINEE GROUP	OCTOBER		FEBRUARY		JUNE	
	Number	Mean	Number	Mean	Number	Mean
All	17,500	314	18,910	312	10,390	317
Male	7,610	314	8,306	313	4,399	317
Female	9,890	313	10,604	311	5,991	316
White	12,010	319	13,230	317	7,117	323
Black	2,170	300	2,028	297	1,033	296
Hispanic	2,216	304	2,412	304	1,555	306
Asian/Pacific Islander	631	304	723	304	395	303
American Indian/AlaskanNative	49	316	56	313	29	321
Non-ResidentAlien	307	300	338	302	188	303
Unknown Race	117	307	123	307	73	314
Community College	9,323	310	10,996	309	7,606	314
State University System	8,177	318	7,914	317	2,784	323

TABLE 16

Mathematics Mean Scaled Scores, 1993-94
(First-Time Examinees in Public Institutions)

EXAMINEE GROUP	OCTOBER		FEBRUARY		JUNE	
	Number	Mean	Number	Mean	Number	Mean
All	17,469	316	18,848	314	10,365	309
Male	7,589	322	8,281	320	4,393	315
Female	9,880	311	10,567	310	5,972	305
White	11,984	320	13,188	319	7,096	313
Black	2,166	298	2,016	297	1,031	291
Hispanic	2,216	307	2,405	305	1,556	301
Asian/Pacific Islander	631	323	722	320	395	316
American Indian/AlaskanNative	48	314	56	307	28	315
Non-ResidentAlien	307	314	338	315	186	315
Unknown Race	117	311	123	310	73	304
Community College	9,302	312	10,940	311	7,579	308
State University System	8,167	320	7,908	319	2,786	313

BIBLIOGRAPHY

American Psychological Association. (1985). *Standards for Educational and Psychological Testing.* Washington, DC: Author.

Brennan, R. L. & Kane, M. T. (1977). An index of dependability for mastery tests. *Journal of Educational Measurement, 14,* 277-298.

Florida Department of Education. *CLAST Test Administration Manual for Institutional Test Administrators, 1986-87.*

Florida Department of Education. *CLAST Test Administration Plan, 1986-87.*

Florida Department of Education. *CLAST Technical Report, 1982-83.*

Florida Department of Education. *CLAST Technical Report, 1983-84.*

Florida Department of Education. (1980). *Procedures for conducting holistic scoring for the essay portion of the College-Level Academic Skills Test.* Tallahassee, FL: Author.

Florida Department of Education. (1981). *Test search and screen for college-level communication and computation skills.* Tallahassee, FL: Author.

Florida Department of Education. (1984). *Test-retest study of the reliability of the College-Level Academic Skills Test.* Tallahassee, FL: Author.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10,* 159-170.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press, 1980.

Ryan, J. ABIL-EST. (Personal communication with J. Ryan of the University of South Carolina, 1981).

Ryan, J. (1981). *Equating new test forms to an existing test.* Paper presented at the Annual Meeting of the National Council of Measurement in Education, Los Angeles.

Stanley, J. C. (1980). Reliability. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd. ed., pp. 356-482). Washington, DC: American Council on Education.

Wright, B. D., Mead, R. J., & Bell, S. R. (1980). *BICAL: calibrating items with the Rasch model.* Research Memorandum No. 23C, Statistical Laboratory, Department of Education, University of Chicago.

APPENDIX A

■ *CLAST Skills Tested, 1993-94*

Essay

- Select a topic which lends itself to development.
- Determine the purpose and the audience for writing.
- Limit the subject to a topic which can be developed within the requirements of time, purpose, and audience.
- Formulate a thesis or main idea statement which focuses the essay.
- Develop the thesis by:
  - providing adequate support which reflects the ability to distinguish between generalized and concrete evidence,
  - arranging the ideas and supporting details in an organizational pattern appropriate to the purpose and focus,
  - writing unified prose in which all supporting material is relevant to the thesis or main idea statement, and
  - writing coherent prose, providing effective transitional devices which clearly reflect the organizational pattern and the relationships of the parts.
- Demonstrate effective word choice.
  - Avoid inappropriate use of slang, jargon, cliches, and pretentious expressions.
- Employ effective sentence structure.
  - Use a variety of sentence patterns.
  - Avoid overuse of passive construction.
- Maintain a consistent point of view.
- Use standard practice for spelling, punctuation, and capitalization.
- Revise, edit, and proofread units of discourse to assure clarity, consistency, and conformity to the conventions of standard American English.

English Language Skills

Word Choice

- Use words which convey the denotative and connotative meanings required by context.
- Avoid wordiness.

Sentence Structure

- Place modifiers correctly.
- Coordinate and subordinate sentence elements according to their relative importance.
- Use parallel expressions for parallel ideas.
- Avoid fragments, comma splices, and fused sentences.

Grammar, Spelling, Capitalization, and Punctuation

- Use standard verb forms.
- Maintain agreement between subject and verb, pronoun and antecedent.
- Use proper case forms.
- Use adjectives and adverbs correctly.

English Language Skills (cont.)

Grammar, Spelling, Capitalization, and Punctuation (cont.)

- Avoid inappropriate shifts in verb tenses.
- Make logical comparisons.
- Use standard practice for spelling, punctuation, and capitalization.

Reading

Literal Comprehension

- Recognize main ideas.
- Identify supporting details.
- Determine the meanings of words on the basis of context.

Critical Comprehension

- Recognize the author's purpose.
- Identify the author's overall organizational pattern.
- Distinguish between statement of fact and statement of opinion.
- Detect bias.
- Recognize the author's tone.
- Recognize explicit and implicit relationships within sentences.
- Recognize explicit and implicit relationships between sentences.
- Recognize valid arguments.
- Draw logical inferences and conclusions.

Mathematics

Arithmetic

- Add and subtract rational numbers.
- Multiply and divide rational numbers.
- Add and subtract rational numbers in decimal form.
- Multiply and divide rational numbers in decimal form.
- Calculate percent increase and percent decrease.
- Solve the sentence "a % of b is c," where values for two of the variables are given.
- Recognize the meaning of exponents.
- Recognize the role of the base number in determining place value in the base-ten numeration system.
- Identify equivalent forms of positive rational numbers involving decimals, percents, and fractions.
- Determine the order-relation between numbers.
- Identify a reasonable estimate of sum, average, or product of numbers.
- Infer relations between numbers in general by examining particular number pairs.
- Solve real-world problems which do not require the use of variables and which do not involve percent.
- Solve real-world problems which do not require the use of variables and which do require the use of percent.
- Solve problems that involve the structure and logic of arithmetic.

Mathematics (cont.)

Algebra

- Add and subtract real numbers.
- Multiply and divide real numbers.
- Apply the order-of-operations agreement to computations involving numbers and variables.
- Use scientific notation in calculations involving very large or very small measurements.
- Solve linear equations.
- Solve linear inequalities.
- Use given formulas to compute results when geometric measurements are not involved.
- Find particular values of a function.
- Factor a quadratic expression.
- Find the roots of a quadratic equation.
- Solve a system of two linear equations in two unknowns.
- Use properties of operations correctly.
- Determine whether a particular number is among the solutions of a given equation or inequality.
- Recognize statements and conditions of proportionality and variation.
- Identify regions of the coordinate plane which correspond to specific conditions and vice versa.
- Use applicable properties to select equivalent equatios or inequalities.
- Solve real-world problems involving the use of variables, aside from commonly used geometric formulas.
- Solve problems that involve the structure and logic of algebra.

Geometry and Measurement

- Round measurements to the nearest given unit of the measuring device.
- Calculate distances, areas, and volumes.
- Identify relationships between angle measures.
- Classify simple plane figures by recognizing their properties.
- Recognize similar triangles and their properties.
- Identify appropriate units of measurement for geometric objects.
- Infer formulas for measuring geometric figures.
- Select applicable formulas for computing measures of geometric figures.
- Solve real-world problems involving perimeters, areas, and volumes of geometric figures.
- Solve real-world problems involving the Pythagorean property.

Logical Reasoning

- Deduce facts of set-inclusion or set non-inclusion from a diagram.
- Identify statements equivalent to the negations of simple and compound statements.
- Determine equivalence or nonequivalence of statements.
- Draw logical conclusions from data.
- Recognize that an argument may not be valid even though its conclusion is true.
- Recognize valid reasoning patterns as illustrated by valid arguments in everyday language.
- Select applicable rules for transforming statements without affecting their meaning.
- Draw logical conclusions when facts warrant them.

Mathematics (cont.)

Statistics, Including Probability

- Identify information contained in bar, line, and circle graphs.
- Determine the mean, median, and mode of a set of numbers.
- Use the fundamental counting principle.
- Recognize properties and interrelationships among the mean, median, and mode in a variety of distributions.
- Choose the most appropriate procedure for selecting an unbiased sample from a target population.
- Identify the probability of a specific outcome in an experiment.
- Infer relations and make accurate predictions from studying statistical data.
- Interpret real-world data involving frequency and cumulative frequency tables.
- Solve real-world problems involving probabilities.

APPENDIX B

■ *College-Level Academic Skills Project (CLASP) and State-Level Task Force Members, 1993-94*

CLASP Task Force Chair

Beth Novinger, Tallahassee Community Collegege

CLASP Staff

Peggy Stillwell, Department of Education
Jana Bielecki, Department of Education

Statewide Test Administrator (STA)

Jeaninne N. Webb, Director
Office of Instructional Resources, University of Florida

Standing Committee on Student Achievement

Robert Stakenas, Chairperson, Florida State University
Linda Adair, Gulf Coast Community College
Laura S. Arnesto, Barry University
Judy Blucker, Florida International University
Maura K. Freeberg, Keiser College
Stephanie Gall, Bay County Schools
Winifred M. Johnson, Bethune-Cookman College
Stuart Lilie, University of Central Florida
Jack McAfee, Indian River County Schools
Lee Rowell, Orange County Schools
Sherry Thomas-Dertke, Pasco-Hernando Community College
Theodore Wright, Broward Community College

Communication Task Force Members

Betsy Griffey, Chairperson, Florida Community College at Jacksonville
Eddie Asgill, Bethune-Cookman College
Jean Bravick, St. Petersburg Junior College
Iris Hart, Santa Fe Community College
Ann Higgins, Gulf Coast Community College
Charles Jimenez, Hillsborough Community College
Patricia Stith, Florida State University
Daniel Strumas, Indian River Community College
Donna Warford-Alley, Daytona Beach Community College
Willa Wolcott, University of Florida

Mathematics Task Force Members

Peter Rautenstrauch, Chairperson, University of Central Florida
Larry Eason, Pasco-Hernando Community College
Emma Fenceroy, Florida A & M University
George Green, Flagler College
Linda Guifoyle, Palmetto High School
Mary Henderson, Okaloosa-Walton Community College
Mary Holton, Central Florida Community College
Gayle Kent, Florida Southern College
Charlene Kincaid, Gulf Breeze High School
Gene Medlin, Stetson University
Judy Miller, Leon High School
Chandra Prayaga, University of West Florida
John Shea, Lake-Sumter Community College
Jaime Villa, Florida Atlantic University

APPENDIX C

■ *Item Review Guidelines*

Overall Factors to Consider in Critiquing Items

1. Adequate measurement of skill

2. Fairness of items: Items should be free of racial, ethnic, sexual, regional, and cultural bias.

3. Quality of stimulus materials (graphics and/or text that provide information required for responding to the item): Content should be
   a. pertinent and appropriate for the grade level,
   b. clear and understandable,
   c. believable and realistic, and
   d. familiar to students of all racial/ethnic backgrounds.

4. Quality of answer choice: The item should contain
   a. one and only one correct answer, neither too obvious and easy nor too difficult and obscure, and
   b. good distractors, neither too obviously incorrect nor too closely related to the correct answer.

5. Readability of items and instructions should follow guidelines set forth in the test item specifications.

6. Quality of language: The language used should be
   a. clear and concise,
   b. appropriate for the grade level,
   c. appropriate for students of all racial/ethnic backgrounds, and
   d. neither too formal and stilted nor too informal and colloquial.

7. Technical considerations: Items should be free from flaws such as
   a. too much variation in the length of response options,
   b. clues in the stem which point to the correct answer,
   c. unclear wording of the stem or directions,
   d. confusing use of negative words in the stem, and
   e. misleading directions–e.g., asking the student to choose the *correct* answer when the *best* answer is really called for (as in choosing the best inference, or the evidence that best supports a given inference), or vice versa.

Questions to Consider in Critiquing Item Construction

1. Stimulus/stem
   a. Does the stem provide ALL THE INFORMATION necessary to answer the question?
   b. Is the desired response evident by reading the stem alone?
   c. Is the stem written in the POSITIVE (avoiding not, except, etc.)?
   d. Is the stimulus portion of the item consistent with the Stimulus Attributes?

2. Response options
    a. Is there an appropriate number of options, arranged in a LOGICAL ORDER (e.g., numerical, alphabetical, chronological)?
    b. Are the options grammatically and conceptually PARALLEL?
    c. Do the options AGREE grammatically with the stem?
    d. Are the options similar and appropriate in LENGTH?
    e. Do the options embody COMMON ERRORS and are they PLAUSIBLE?
    f. Do the options AVOID "all of the above" or "none of the above"?

3. The entire item
    a. Does the item avoid tricky words, phrases, and constructions?
    b. Is the item free of superfluous material and awkward wording?
    c. Does the item avoid unnecessary clues?
    d. Does the item focus on IMPORTANT aspects of content, not trivia?

Considerations in Critiquing Items for Bias

An item is considered to be biased if it contains any language or vocabulary that could benefit or hinder any group's performance. When reviewing an item for bias, one must consider all of the following groups:

females	regional groups within the U.S.
males	international groups
racial/ethnic groups	religious groups
cultural groups	the visually impaired
age groups	the hearing impaired
socio-economic groups	persons with other handicaps

As you review each item, consider each of the following questions:

1. Does the item contain any information that could seem offensive to any group?

2. Does the item include or imply any stereotypic depiction of any group?

3. Does the item portray any group as degraded in any way?

4. Does the item contain any group-specific language or vocabulary (e.g., culture-related expressions, slang, or expressions) that may be unfamiliar to particular examinees?

APPENDIX D

■ *Test-Retest Reliability of the CLAST*

In 1984, the Department of Education contracted with Dr. F. J. King of the Florida State University to study certain aspects of the reliability of the College-Level Academic Skills Test (CLAST). Dr. King prepared a report entitled "A Test-Retest Study of the Reliability of the College-Level Academic Skills Test." The study is available from the Department of Education and is summarized herein.

Dr. King invited 360 students who had taken the CLAST in September 1984 to take the CLAST examination a second time. Two hundred seventy-four agreed to do so, and 220 usable scores were obtained. The students were retested in October 1984 with the same form of the test which had been administered in June 1984.

The data were analyzed using several statistics. A Hambleton-Novick (1973) index was calculated to obtain an estimate of the decision consistency over two test forms. The Brennan-Kane (1977) index was used to obtain an index of decision consistency for a single test administration. The KR-20 (Stanley, 1971) index was also calculated because it is a reliability coefficient widely used with norm-referenced tests.

The Hambleton-Novick index calculated with the 1984 passing criteria resulted in the following:

Computation	0.97
Reading	0.86
Writing	0.96
Essay	0.86

The Brennan-Kane indices for the subtests were as follows:

Computation	0.96
Reading	0.96
Writing	0.92
Essay	not applicable

The KR-20 internal consistency coefficients for the subtests resulted in values of:

Computation	0.83
Reading	0.87
Writing	0.74
Essay	not applicable

The reliability coefficients varied depending on which test administration was being analyzed, the relative difficulty of the tests, and the psychometric characteristics of the tests themselves. Further, it must be recognized that the reported reliability coefficients will vary for subpopulations (e.g., Hispanic) and will vary depending on the placement of the passing criterion.

FLORIDA DEPARTMENT OF
EDUCATION

State of Florida
Department of Education
College-Level Academic Skills Project
Tallahassee, Florida
Doug Jamerson, Commissioner
Affirmative action/equal opportunity employer

197-092394-450-JB-1